

日本国特許庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されて
いる事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed
with this Office

出願年月日

Date of Application:

2001年11月28日

出願番号

Application Number:

特願2001-362858

ST.10/C]:

[JP2001-362858]

出願人

Applicant(s):

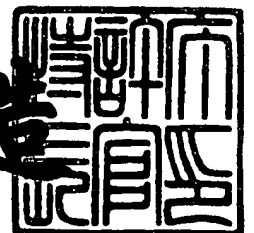
株式会社日立製作所

CERTIFIED COPY OF
PRIORITY DOCUMENT

2002年 2月 5日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造



【書類名】 特許願

【整理番号】 KN1350

【提出日】 平成13年11月28日

【あて先】 特許庁長官殿

【国際特許分類】 G06F 13/00

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社
 日立製作所 システム開発研究所内

 【氏名】 水野 陽一

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社
 日立製作所 システム開発研究所内

 【氏名】 松並 直人

【発明者】

 【住所又は居所】 神奈川県川崎市麻生区王禅寺 1 0 9 9 番地 株式会社
 日立製作所 システム開発研究所内

 【氏名】 味松 康行

【発明者】

 【住所又は居所】 神奈川県小田原市中里 3 2 2 番地 2 号 株式会社 日立
 製作所 R A I D システム事業部内

 【氏名】 ▲高▼本 賢一

【特許出願人】

 【識別番号】 000005108

 【氏名又は名称】 株式会社 日立製作所

【代理人】

 【識別番号】 100093492

 【弁理士】

 【氏名又は名称】 鈴木 市郎

 【電話番号】 03-3591-8550

【選任した代理人】

【識別番号】 100078134

【弁理士】

【氏名又は名称】 武 顕次郎

【手数料の表示】

【予納台帳番号】 113584

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 ディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法

【特許請求の範囲】

【請求項 1】 1 または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用されるディスクアレイ装置とにより構成されるディスクアレイシステムにおいて、前記コントローラは、ディスク装置に作成される論理ユニットの構成情報を管理する構成管理手段を備え、該構成管理手段は、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことを特徴とするディスクアレイシステム。

【請求項 2】 前記コントローラ間での論理ユニットの引き継ぎの際、対象論理ユニットと移行先コントローラとが指定されることを特徴とする請求項 1 記載のディスクアレイシステム。

【請求項 3】 前記コントローラ間での論理ユニットの引き継ぎの際、移行先コントローラでの接続ポート番号または前記計算機から認識される論理ユニット番号が変更されることを特徴とする請求項 1 記載のディスクアレイシステム。

【請求項 4】 前記コントローラ間での論理ユニットの引き継ぎの際、前記キャッシュ上の対象論理ユニットのデータを前記ディスク装置に書き込むことを特徴とする請求項 1 記載のディスクアレイシステム。

【請求項 5】 前記キャッシュ上のデータの書き込みは、書き込みを指示するコマンドによって行われることを特徴とする請求項 4 記載のディスクアレイシステム。

【請求項 6】 前記構成管理手段は、移行先コントローラを記憶しておくことにより、自動的な論理ユニットの引き継ぎを行うことを特徴とする請求項 1 記載のディスクアレイシステム。

【請求項 7】 1 または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用

されるディスクアレイ装置とにより構成されるディスクアレイシステムのコントローラ間での論理ユニットの引き継ぎ方法において、前記コントローラは、ディスク装置に作成される論理ユニットの構成情報を管理し、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことを特徴とするコントローラ間での論理ユニットの引き継ぎ方法。

【請求項 8】 前記コントローラ間での論理ユニットの引き継ぎの際、前記キャッシュ上の対象論理ユニットのデータを前記ディスク装置に書き込むことを特徴とする請求項 7 記載のコントローラ間での論理ユニットの引き継ぎ方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、ディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法に係り、特に、複数のコントローラのそれぞれが専用のキャッシュを備えたディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法に関する。

【0002】

【従来の技術】

ディスクアレイシステムに関する従来技術として、複数のディスク装置のそれぞれに対応して設けられたコントローラと、各コントローラからアクセス可能な共有キャッシュとを備えたディスクアレイシステムが知られている。

【0003】

図 10 は共有キャッシュを備えた従来技術によるディスクアレイシステムの構成を示すブロック図であり、図 10 を参照して従来技術について説明する。図 10 において、100 はディスクアレイ装置、200x (x = a …… n) はコントローラ、300 は共有キャッシュ、400 は共有メモリ、500x (x = a …… n) はディスク装置、600 は共通バスである。

【0004】

図 10 に示すディスクアレイ装置 100 は、図示しない他の計算機等と共に、

ディスクアレイシステムを構成するものであり、ディスク装置を説明する複数のコントローラ 2 0 0 x と、入出力データを格納する共有キャッシュ 3 0 0 と、構成情報等を格納する共有メモリ 4 0 0 と、複数のディスク装置 5 0 0 x と、これらを接続する共通バス 6 0 0 とにより構成されている。共有キャッシュ 3 0 0 及び共有メモリ 4 0 0 は、どのコントローラ 2 0 0 x から共通バス 6 0 0 を通してアクセス可能である。共有キャッシュ 3 0 0 は、全てのコントローラ 2 0 0 x からアクセスが集中するため、高価な大容量のものが用意される。そして、各コントローラ 2 0 0 x は、それぞれ担当するディスク装置 5 0 0 x が予め決められている。

【 0 0 0 5 】

前述した構成の従来技術は、あるコントローラ 2 0 0 x が担当するディスク装置 5 0 0 x を別のコントローラ 2 0 0 x に切り替える場合、ディスク装置 5 0 0 x に関する入出力データが共有キャッシュ 3 0 0 上に置かれているため、移行先のコントローラが共有キャッシュ上のデータをそのまま使用して、ディスク装置 5 0 0 x とコントローラ 2 0 0 x との関係を変更することができる。また、引継ぎに必要なディスク装置の構成情報等は共有メモリ 4 0 0 に格納されており、移行先コントローラは、共有メモリ 4 0 0 を参照して、対象ディスク装置の制御を開始すればよい。また、移行元コントローラは、これまで制御していたディスク装置 5 0 0 x に対するアクセス要求を拒否し、ディスク装置 5 0 0 x の管理を止めるだけでよい。これにより、前述の共有キャッシュを備えた従来技術によるディスクアレイシステムは、あるコントローラが担当するディスク装置を切り替えることにより、コントローラが担当するディスク装置内の論理ユニットに対するアクセス量を均等化してコントローラの負荷分散を行うことができる。

【 0 0 0 6 】

また、前述の従来技術は、ディスクを移設し、物理位置を移動することによって、負荷分散を図ることが可能である。この場合、コントローラは、対象となるディスク装置へのアクセスを停止し、ディスク装置をシステムから完全に切り離し、次に、切り離れたディスク装置が物理的に異なる位置に差し込まれたとき、ディスク装置を再認識して処理を開始する。また、前述の従来技術は、ディスク

を異なるディスクアレイ装置に移動することも可能である。これにより、前述の従来技術は、再認識されたディスク装置を新しい別のコントローラに割り当て、制御するコントローラを切替えることにより、前述と同様に、負荷分散を行うことができる。

【 0 0 0 7 】

【発明が解決しようとする課題】

前述した従来技術は、共有キャッシュを備えたディスクアレイ装置に適用した場合に好適なものであったが、各コントローラが独立した専用キャッシュを備えたディスクアレイ装置に適用することができなかった。

【 0 0 0 8 】

各コントローラが独立した専用キャッシュを備えている分散キャッシュ環境のディスクアレイシステムは、各コントローラがそれぞれ専用キャッシュを使用して個々のボリュームを担当することにより、共有キャッシュ環境の場合に生じていたようなアクセスの集中を回避し、コストパフォーマンス及びスケーラビリティを上げることができる。

【 0 0 0 9 】

しかし、各コントローラが独立した専用キャッシュを備えている分散キャッシュ環境のディスクアレイシステムは、あるコントローラが担当するディスク装置を切り替えようとする場合、移行先コントローラが対象となるボリュームの入出力データを保持していないため、コントローラをそのまま切り替えることができないという問題点を有している。

【 0 0 1 0 】

このような問題は、全てのコントローラのデータキャッシュに常に同じ内容を書き込んでおけば解決することができ、共有キャッシュの場合と同様にボリュームを引き継ぐようにすることが可能であるが、各コントローラのキャッシュに担当でないボリュームのデータも書き込むことになり、キャッシュ容量を圧迫し、分散キャッシュのメリットが失われることになるという問題点を生じる。

【 0 0 1 1 】

また、各コントローラが独立した専用キャッシュを備えているディスクアレイ

システムは、ディスクの移設の自動的な切り替え及びダイナミックな切り替えを行うことができないという問題点を有している。すなわち、分散キャッシュ環境のディスクアレイシステムは、ディスクをシステムから切り離す作業と、切り離れたディスクの物理位置を変更する作業と、ディスクを再認識させる作業とが必要であり、人手を介してこれらの作業を行わなければならない、この間、対象ディスクへのアクセスを完全に停止しなければならない、しかも、この場合、ディスクを元のコントローラに戻すことは考慮されていなかった。

【 0 0 1 2 】

前述したように、各コントローラが独立した専用キャッシュを備えている従来技術のディスクアレイシステムは、システムを停止せずにダイナミックに担当コントローラを切り替えることができないという問題点を有している。

【 0 0 1 3 】

本発明の目的は、前述した従来技術の問題点を解決し、システムを中断させることなく任意のコントローラ間で任意のボリュームをダイナミックに引継ぐことを可能とした、各コントローラが独立した専用キャッシュを備え、それぞれ個々のボリュームを担当しているディスクアレイシステム及びコントローラ間での論理ユニットの引き継ぎ方法を提供することにある。

【 0 0 1 4 】

【課題を解決するための手段】

本発明によれば前記目的は、1または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使用されるディスクアレイ装置とにより構成されるディスクアレイシステムにおいて、前記コントローラが、ディスク装置に作成される論理ユニットの構成情報を管理する構成管理手段を備え、該構成管理手段は、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことにより達成される。

【 0 0 1 5 】

また、前記目的は、1または複数の計算機と、それぞれが専用のキャッシュを備えた複数のコントローラ及び複数のディスク装置を備えた前記計算機により使

用されるディスクアレイ装置とにより構成されるディスクアレイシステムのコントローラ間での論理ユニットの引き継ぎ方法において、前記コントローラが、ディスク装置に作成される論理ユニットの構成情報を管理し、前記論理ユニットの構成情報を書き替えることにより前記コントローラが担当する任意の論理ユニットを別の任意のコントローラに引き継ぐ処理を行うことにより達成される。

【 0 0 1 6 】

【発明の実施の形態】

以下、本発明によるディスクアレイ装置の実施形態を図面により詳細に説明する。

【 0 0 1 7 】

図 1 は本発明の第 1 の実施形態によるディスクアレイシステムの構成を示すブロック図、図 2 は構成管理手段が備える構成情報テーブルの構成例を示す図である。図 1、図 2 において、1 はディスクアレイ装置、 $2x$ ($x = a, b, \dots, n$) は計算機、3 は経路制御装置、4 は管理コンソール、 $5x$ ($x = a, b, \dots, n$) はチャネルパス、 $6x$ ($x = a, b, \dots, n$) は LAN (Local Area Network)、7 は通信手段、10 はデバイスネットワーク、 $11x$ ($x = a, b, \dots, n$) はコントローラ、 $12x$ ($x = a, b, \dots, n$) はディスク装置、31 はパス変更手段、41 は管理ユーティリティ、 $111x$ ($x = a, b, \dots, n$) はデータキャッシュ、 $112x$ ($x = a, b, \dots, n$) は構成管理手段、 $121a \sim 121f$ は LU (Logical Unit) である。

【 0 0 1 8 】

本発明の第 1 の実施形態は、図 1 に示すように、ディスクアレイ装置 1 と、ディスクアレイ装置 1 を使用する複数の計算機 $2x$ と、全ての計算機 $2x$ とディスクアレイ装置 1 と管理コンソール 4 を相互に接続する経路制御装置 3 と、ディスクアレイ装置 1 を管理するための管理コンソール 4 とが、チャネルパス $5x$ により相互に接続されて構成されている。図示例では、チャネルパス $5x$ して、ファイバチャネルを使用することとする。また、図示実施形態は、計算機 $2x$ と管理コンソール 4 とが通信を行うための LAN $6x$ と、ディスクアレイ装置 1 と管理コンソール 4 とが通信するための通信手段 7 とが備えられる。

【 0 0 1 9 】

経路制御装置 3 には、計算機 2 x からディスクアレイ装置 1 へのアクセスパスを変更するためのプログラムであるパス変更手段が備えられる。図 1 に示す例では、パス変更手段 3 1 を経路制御装置 3 上に配置するものとしているが、同様の機能を計算機 2 x 上に配置するようにすることもできる。

【 0 0 2 0 】

管理コンソール 4 には、ディスクアレイ装置 1 内のディスク装置 1 2 x 内部の L U 構成の表示や、システムの管理者がディスクアレイ装置 1 の L U 設定等を行うために使用される管理ユーティリティ 4 1 が備えられる。図 1 に示す例では、管理ユーティリティ 4 1 を管理コンソール 4 上に配置するものとしているが、管理ユーティリティ 4 1 を計算機 2 x や、ディスクアレイ装置 1 に配置するようにすることもできる。

【 0 0 2 1 】

ディスクアレイ装置 1 は、複数のコントローラ 1 1 x と、複数のディスク装置 1 2 x とが、デバイスネットワーク 1 0 で接続されて構成される。この構成により、任意のコントローラから任意のディスク装置へのアクセスが可能である。デバイスネットワーク 1 0 は、どのようなインタフェースによって構成してもよいが、接続性に優れたファイバチャネル等の利用が好適であり、また、総合的なスループットを高めるため、1 つあるいは複数のスイッチで構成することも可能である。

【 0 0 2 2 】

コントローラ 1 1 x は、計算機 2 から書き込まれたデータや、ディスク装置 1 2 x から読み出されたデータを一時的に格納する入出力データ格納用のデータキャッシュ 1 1 1 x と、ディスクや L U の構成を管理する構成管理手段 1 1 2 x とを備えて構成される。ここで、L U (Logical Unit) とは、ディスクアレイ装置 1 内部に設けた仮想的な論理ボリュームのことであり、計算機 2 とディスクアレイ装置 1 とを接続するインターフェイスの 1 つのプロトコルである S C S I (Small Computer System Interface) の仕様において定義された名称である。また、L U を識別するための番号のことを論理ボリューム番号 (L U N : Logical Unit

Number)と呼ぶ。

【 0 0 2 3 】

ディスクアレイ装置 1 の内部で定義した LU を特に内部論理ボリューム（内部 LU）と呼び、ディスクアレイ装置 1 は、内部 LU を管理するため、LU に 0 から始まる整数でシリアル番号を付ける。この番号を内部論理ボリューム番号（内部 LUN）と呼ぶ。計算機 2 x は、LU を検出する際に LUN を 0 から順にサーチし、ある番号が存在しない場合それ以降のサーチを行わない場合がある。このため、内部 LUN をそのまま計算機 2 x に割り当てるのではなく、計算機 2 x が認識できる LUN に変換して割り当てる必要がある。このようにして各計算機 2 x から認識される LUN を外部論理ボリューム番号（外部 LUN）と呼び、内部 LUN とは区別される。

【 0 0 2 4 】

論理ボリューム LU 1 2 1 x（1 2 1 a、1 2 1 b、……、1 2 1 n）は、それぞれディスク装置 1 2 x 上に作成される。なお、LU 1 2 1 x は、どのような RAID 構成であってもよい。論理ボリューム LU の作成時に、該当する LU の制御を担当するコントローラが割り当てられる。1 つのコントローラに対して複数の LU を割り当てることも可能である。

【 0 0 2 5 】

図 1 に示す構成を有するディスクアレイシステムは、各コントローラ 1 1 x が専用のデータキャッシュ 1 1 1 x を備え、個々の LU の制御を担当することにより、特定のデータキャッシュへのアクセスの集中を防止することができ、共有キャッシュを備えるシステムに比べてコストパフォーマンス及びスケーラビリティを高めることができる。

【 0 0 2 6 】

構成管理手段 1 1 2 x が備える構成情報テーブル 1 1 2 1 x の一例を図 2 に示している。構成情報テーブル 1 1 2 1 x は、各 LU 1 2 1 x の構成情報を管理するためのテーブルであり、内部 LUN、外部 LUN、ポート番号、コントローラ番号、ブロック数、RAID グループ番号、RAID レベル、物理アドレス情報等が格納される。なお、この構成情報テーブル 1 1 2 1 x は、ディスク装置 1 2

xの予め定められた位置に格納されていてよい。

【0027】

前述において、ポート番号は、コントローラ11xが備えるファイバチャネル接続ポートのうち、そのLUが使用できるポートの識別番号を表す。コントローラ番号は、ディスクアレイ装置1内のコントローラの識別番号であり、その中のデフォルトコントローラ番号は、本来どのコントローラがそのLUを担当すべきかを表し、カレントコントローラ番号は、現在制御を担当しているコントローラを表す。ブロック数は、各LUの論理ブロック数を表し、これにより、各LUのサイズを知ることができる。

【0028】

ディスクアレイ装置1は、その内部で、同一RAIDレベルのLUをグループ化して管理している場合がある。RAIDグループ番号、RAIDレベルは、その場合の各LUが属するRAIDグループの識別番号及びそのRAIDレベルを表す。物理アドレス情報は、各LUの論理アドレスに対する物理的なディスク位置情報である。

【0029】

図3はコントローラ相互間でLUの引き継ぎを行う場合の処理動作を説明するフローチャートであり、次に、コントローラ間のLUの引き継ぎについて説明する。

【0030】

(1) 管理コンソール4を操作する作業者は、移行対象となる全LUのLUNと、移行先コントローラとを指定する。その際、移行先コントローラで使用するポート番号、外部LUNを変更したい場合は、それぞれポート番号、外部LUNも併せて指定する(ステップ801)。

【0031】

(2) 管理ユーティリティ41は、指定された情報に基づいて通信手段7を介してディスクアレイ装置1にLUの引き継ぎ指示を発行する(ステップ802)。

【0032】

(3) 移行元コントローラ11xの構成管理手段112xは、LUの引き継ぎ指

示を受け取ると、対象LUの構成情報を移行先コントローラ11xへ移行する。具体的には、移行元コントローラ11xの構成管理手段112xが構成情報テーブル1121xの対象LUの欄をディスクの予め決められた位置へ書き込み、移行先コントローラ11xの構成管理手段112xがそれを読み込むことにより、対象LUの構成情報を移行する。ポート番号、外部LUNを変更する場合、移行元コントローラ11xの構成管理手段112xが、構成情報テーブル1121xを変更し、変更後の内容をディスクに書き込む。構成情報の移行は、デバイスネットワーク10を経由して送信することにより行われる。また、図示しない専用線を経由で直接送信するようにすることも可能である。また、全コントローラ11xの構成管理手段112xが、予め全LUの構成情報を構成情報テーブル1121xに格納しておくことももちろん可能である。その場合、構成情報の移行を省略することができ、ポート番号、外部LUN、コントローラ番号のみを変更すればよい（ステップ803）。

【0033】

（4）次に、移行元コントローラ11xは、データキャッシュ111x上にある対象LUのデータを全てデステージする。デステージとは、データキャッシュ上のデータをディスクに書きこむ処理のことである。この間、計算機2xからのライトアクセスは全てライトスルーとする。これにより、対象LUのデータをデータキャッシュ上から全て掃き出し、ディスク上のLUの内容を整合性が保たれた状態にすることができる。なお、デステージの処理は、管理コンソール4からの書き込みコマンドに基づいて行うようにすることもできる（ステップ804）。

【0034】

（5）移行元コントローラ11xは、対象LUのデステージが完了すると、管理ユーティリティ41にデステージ完了を通知し、デステージ完了の通知を受けた管理ユーティリティ41は、パス変更手段31へパス切り替えを指示する（ステップ805）。

【0035】

（6）パス切り替えの指示を受けたパス変更手段31は、対象LUのアクセスを一時保留する。これはLU引継ぎ中の過渡状態でのアクセスを抑制するためであ

る（ステップ 8 0 6）。

【 0 0 3 6 】

（ 7 ）次に、パス変更手段 3 1 は、対象 LU のアクセスパスを移行元コントローラから移行先コントローラへ切り替える。これにより、対象 LU へのフレームは、全て移行先コントローラへ送信されることになる（ステップ 8 0 7）。

【 0 0 3 7 】

（ 8 ）次に、対象 LU の制御を行うコントローラを移行先コントローラに切り替える。移行元コントローラは、対象 LU に対するアクセスを完全に停止し、移行先コントローラは対象 LU の制御を開始する（ステップ 8 0 8）。

【 0 0 3 8 】

（ 9 ）コントローラの切り替えが完了すると、パス変更手段 3 1 は、対象 LU のアクセスを再開させる（ステップ 8 0 9）。

【 0 0 3 9 】

前述した処理により、LU の引き継ぎが完了する。前述のような処理により、キャッシュ上の対象 LU のデータをデステージして、ディスクの整合性を保つことができ、各コントローラが独立した専用キャッシュを備えた場合においても、システムを中断させることなく任意のコントローラ間でボリュームを引継ぐことができる。

【 0 0 4 0 】

図 4 は図 3 により説明した LU 引き継ぎの具体的な例を説明する図、図 5 は引き継ぎ前後の構成テーブルの例を示す図であり、次に、図 4、図 5 を参照して、LU 引き継ぎの具体例と構成情報テーブルの書き換えとについて説明する。

【 0 0 4 1 】

いま、ある計算機 2 x が、図 4 （ a ）に示すように、コントローラ 0[#] 1 1 a に割り当てられている LU 0[#] 1 2 1 a ～ LU 3[#] 1 2 1 d を使用しているものとする。そして、コントローラ 1[#] 1 1 b を追加し、図 4 （ b ）に示すように、LU 2[#] 1 2 1 c、LU 3[#] 1 2 1 d の制御をコントローラ 1[#] 1 1 b に切り替えることとする。このとき、外部 LUN は変更せず、ポート番号のみを変更するものとする。また、簡単のため、全構成管理手段 1 1 2 x が全 LU の情報を構成

情報テーブル 1 1 2 1 x に格納しているものとする。

【 0 0 4 2 】

前述の場合、引き継ぎ前の構成情報テーブル 1 1 2 1 x は、図 5 (a) に示すような状態となっている。なお、ブロック数、RAID グループ番号、RAID レベル、物理アドレス情報等は引き継ぎによって変更されることはないので、ここでは図示を省略している。

【 0 0 4 3 】

図 5 (a) から全ての LU に対して、デフォルトコントローラ番号及びカレントコントローラ番号に、現在の担当コントローラである“0”が格納されており、また、LU 0 及び LU 1 は、ポート番号 0 を、LU 2 及び LU 3 はポート番号 1 を使用していることが判る。

【 0 0 4 4 】

管理コンソール 4 を操作する作業者は、移行対象である LU 2[#] 1 2 1 c、LU 3[#] 1 2 1 d を選択して、移行先のコントローラ 1[#] 1 1 b 及び移行先コントローラ 1[#] 1 1 b で使用するポート番号“0”を指定する。指定された情報に基づいて、構成情報テーブル 1 1 2 1 x が変更される。この結果、変更後の構成情報テーブル 1 1 2 1 x は、図 5 (b) に示すような状態になる。図 5 (b) から LU 2[#] 及び LU 3[#] のポート番号が“0”に、デフォルトコントローラ番号及びカレントコントローラ番号が“1”に変更されたことが判る。これにより、LU 2[#] 1 2 1 c、LU 3[#] 1 2 1 d の制御をコントローラ 1[#] 1 1 b に切り替えることができる。

【 0 0 4 5 】

前述した本発明の第 1 の実施形態によれば、データキャッシュ上の対象ボリュームのデータをディスクに掃き出し、ディスクの整合性を保つことにより、各コントローラが独立した専用キャッシュを備えた場合においても、システムを中断させることなく任意のコントローラ間で任意のボリュームを引き継ぐことができ、システムにおける各コントローラの負荷分散を図ることができるという効果を得ることができる。

【 0 0 4 6 】

図 6 は本発明の第 2 の実施形態によるディスクアレイシステムの構成を示すブロック図である。図 6 において、20 はテープ装置であり、他の符号は図 1 の場合と同一である。

【0047】

図 6 に示す本発明の第 2 の実施形態における図 1 に示した第 1 の実施形態との相違点は、計算機 2n にテープ装置 20 が接続されていることと、各計算機 2x が各コントローラ 11x に直接接続されていることとである。もちろん、第 1 の実施形態の場合と同様に、計算機 2x とコントローラ 11x とを経路制御装置を介して接続してもよいが、第 2 の実施形態は、パス変更手段を不要とすることができる。また、計算機 2x とコントローラ 11x とは、1 対 1 に対応している必要はなく、1 つのコントローラに対して、複数の計算機を接続することも可能である。

【0048】

図 6 に示す本発明の第 2 の実施形態は、データ二重化機能を備え、計算機 2n を、ディスクアレイ装置 1 内のボリュームのバックアップを取得するために使用する一般にバックアップサーバと呼ぶ計算機としたものである。本発明の第 2 の実施形態は、バックアップ処理に限らず、バッチ処理やデータマイニング等の様々な処理に適用可能であるが、ここでは、本発明をバックアップ処理に適用する場合を想定して説明する。

【0049】

図 6 に示すディスクアレイシステムにおいて、コントローラ 11a には、LU0[#] 121g 及び LU1[#] 121h が割り当てられ、コントローラ 11b には、LU2[#] 121i 及び LU3[#] 121j が割り当てられているものとする。そして、LU1[#] 121h、LU3[#] 121j は、それぞれ、LU0[#] 121g、LU2[#] 121i の複製であり、LU0[#] 121g、LU2[#] 121i と同一の内容のデータが格納されている。このようにディスクアレイ装置内にボリュームの複製を作成する機能を一般にデータ二重化機能と呼ぶ。

【0050】

図 7 はデータ二重化機能を利用したコントローラ相互間で LU の引き継ぎを行

う場合の処理動作を説明するフローチャートであり、次に、これについて説明する。ここで説明する例は、バックアップサーバとしての計算機 n が、二重化されている LU の内容をテープ装置 20 内にバックアップする場合の LU の引き継ぎの例である。

【0051】

(1) まず、計算機 $2x$ は、バックアップ対象 LU に対するペア分割指示をディスクアレイ装置 1 に送る (ステップ 901)。

【0052】

(2) ペアが分割されると、分割された LU を担当している移行元コントローラ $11x$ の構成管理手段 $112x$ は、構成情報テーブル $1121x$ のデフォルトコントローラ番号で指定されたコントローラ $11n$ へ、対象 LU の構成情報を移行する。移行の際は、デフォルトコントローラ番号は、移行元コントローラの番号に変更される。また、構成情報の他に、データ二重化機能が使用する LU のステータスや、LU の書き込み位置を記録する差分情報等の必要な管理情報も併せて移行する。移行方法は、前述した第 1 の実施形態の場合と同様に任意の方法により行うことができる (ステップ 902)。

【0053】

(3) 次に、移行元コントローラ $11x$ は、データキャッシュ $111x$ 上の対象 LU のデータを全てデステージする。この間、計算機 $2x$ からのライトアクセスは全てライトスルーとし、対象 LU のデータをデータキャッシュ $111x$ 上から全て掃き出す。その際、計算機 $2x$ から対象 LU のデステージを指示するコマンドを発行してもよい (ステップ 903)。

【0054】

(4) デステージが完了すると、移行元コントローラ $11x$ は、移行先コントローラへデステージの完了を通知し、対象 LU の制御を移行先コントローラ $11n$ に切り替える (ステップ 904)。

【0055】

(5) 対象 LU の制御が切り替わり、移行先コントローラ $11n$ から対象 LU へのアクセスが可能になると、計算機 $2n$ は、対象 LU のデータをテープ装置 20

等に格納してバックアップを取得する（ステップ905）。

【0056】

（6）バックアップの取得が完了すると、計算機2nは、対象LUに対するペア再同期指示をディスクアレイ装置1に送る（ステップ906）。

【0057】

（7）ペア再同期指示を受けると、ステップ902の処理と同様に、対象LUの制御を担当するコントローラ11nの構成管理手段112nは、構成情報テーブル1121nのデフォルトコントローラ番号で指定されたコントローラ11xへ、対象LUの構成情報を移行する。移行の際、デフォルトコントローラ番号は、コントローラ11nの番号に変更される。また、構成情報の他に、データ二重化機能が使用するLUのステータスや、LUの書き込み位置を記録する差分情報等の必要な管理情報も併せて移行する（ステップ907）。

【0058】

（8）次に、ステップ903の処理と同様に、対象LUの制御を担当するコントローラ11nは、データキャッシュ111n上の対象LUのデータを全てデステージする。この間、計算機2nからのライトアクセスは全てライトスルーとし、対象LUのデータをデータキャッシュ111n上から全て掃き出す（ステップ908）。

【0059】

（9）デステージが完了すると、対象LUの制御を担当するコントローラ11nは、デフォルトコントローラ番号で指定されたコントローラ11xへデステージの完了を通知し、対象LUの制御を元のコントローラ11xへ戻して処理を終了する（ステップ909）。

【0060】

前述した本発明の第2の実施形態によれば、バックアップ処理を専用のコントローラで行うようにすることにより、通常運用系の性能に影響を与えることなくデータのバックアップを行うことが可能である。

【0061】

図8は図7により説明したLU引き継ぎの具体的な例を説明する図、図9は引

き継ぎ前後の構成テーブルの例を示す図であり、次に、図 8、図 9 を参照して、LU 引き継ぎの具体例と構成情報テーブルの書き換えとについて説明する。

【 0 0 6 2 】

いま、図 8 (a) に示すように、コントローラ $0^{\#} 11a$ に $LU0^{\#} 121g$ 、 $LU1^{\#} 121h$ が割り当てられており、コントローラ $1^{\#} 11b$ に $LU2^{\#} 121i$ 、 $LU3^{\#} 121j$ が割り当てられているものとする。そして、図 8 (b) に示すように、 $LU1^{\#} 121h$ 及び $LU3^{\#} 121j$ の制御をコントローラ $n^{\#} 11n$ に切り替えるものとする。なお、ここで説明している例では、外部 LUN 及びポート番号は変更せずそのまま使用できるものとし、また、第 1 の実施形態の場合と同様に、全構成管理手段 $112x$ が全 LU の情報を構成情報テーブル $1121x$ に格納しているものとする。

【 0 0 6 3 】

前述の場合、引き継ぎ前の構成情報テーブル $1121x$ は、図 9 (a) に示すような状態となっている。この図から、カレントコントローラ番号には、 $LU0$ 及び $LU1$ に対して “0” が、 $LU2$ 及び $LU3$ に対して “1” が格納されており、現在の制御を担当しているコントローラ番号を表していることがわかる。そして、 $LU1$ 及び $LU3$ のデフォルトコントローラ番号は移行先コントローラである n が格納されている。

【 0 0 6 4 】

図 9 (a) に示す構成情報テーブルから移行元であるコントローラ $0^{\#} 11a$ の構成管理手段 $112a$ 及びコントローラ $1^{\#} 11b$ の構成管理手段 $112b$ は、デフォルトコントローラ番号で指定されているコントローラ $n^{\#} 11n$ が移行先であることを知ることができる。移行の際、構成情報テーブル $1121x$ のデフォルトコントローラ番号は移行元コントローラの番号に変更される。この結果、引き継ぎ後の構成情報テーブルは、図 9 (b) に示すようになる。このテーブルから $LU1$ 及び $LU3$ のデフォルトコントローラ番号がそれぞれ “0” 及び “1” に変更されていることが判る。また、引き継ぎ後の制御を担当するコントローラはコントローラ n であり、カレントコントローラ番号が n に変更されている。

【 0 0 6 5 】

コントローラを切り替えてバックアップ処理が終了すると、1回目の移行とは逆向きのコントローラの移行が行われる。この場合、1回目と同様に、コントローラ $n^{\#} 1 1 n$ の構成管理手段 $1 1 2 n$ は、構成情報テーブル $1 1 2 1 n$ のデフォルトコントローラ番号から移行先コントローラを知ることができる。これにより、自動的にLUの制御を元のコントローラに戻すことができる。LUの制御を戻したときのコントローラとLUの割り当てとは図8(a)に示す場合と同等である。対象LUの制御を戻す際、構成情報テーブル $1 1 2 1 x$ の対象LUのデフォルトコントローラ番号は n に変更される。このときの構成情報テーブル $1 1 2 1 x$ は、図9(a)に示す場合と同等であり、LU2及びLU3のデフォルトコントローラ番号が再び n に変更されていることが判る。前述したようにして繰り返しLUの引継ぎを行うことができる。

【0066】

前述した本発明の第2の実施形態によれば、バックアップ処理を専用のコントローラで行うようにすることにより、通常運用系の性能に影響を与えることなくバックアップを行うことが可能となる。また、構成情報テーブルに移行すべきコントローラ番号を記録しておくことにより、自動的にLUの引き継ぎを行うことが可能となり、また、自動的に元のコントローラに制御を戻して、通常の処理を継続して行うことができる。

【0067】

【発明の効果】

以上説明したように本発明によれば、各コントローラが独立した専用キャッシュを備えた場合においてもシステムを中断させることなく任意のコントローラ間で任意のボリュームを引継ぐことができ、システムにおける各コントローラの負荷分散を図ることができるという効果を奏することができる。

【0068】

また、本発明によれば、自動的なボリュームの引継ぎ及び元のコントローラに制御を戻すことができ、これにより、バックアップ処理を専用のコントローラで行うことができ、通常運用系の性能に影響を与えることなくバックアップを行うことができる。

【図面の簡単な説明】

【図 1】

本発明の第 1 の実施形態によるディスクアレイシステムの構成を示すブロック図である。

【図 2】

構成管理手段が備える構成情報テーブルの構成例を示す図である。

【図 3】

コントローラ相互間で LU の引き継ぎを行う場合の処理動作を説明するフローチャートである。

【図 4】

図 3 により説明した LU 引き継ぎの具体的な例を説明する図である。

【図 5】

引き継ぎ前後の構成テーブルの例を示す図である。

【図 6】

本発明の第 2 の実施形態によるディスクアレイシステムの構成を示すブロック図である。

【図 7】

データ二重化機能を利用したコントローラ相互間で LU の引き継ぎを行う場合の処理動作を説明するフローチャートである。

【図 8】

図 7 により説明した LU 引き継ぎの具体的な例を説明する図である。

【図 9】

引き継ぎ前後の構成テーブルの例を示す図である。

【図 10】

共有キャッシュを備えた従来技術によるディスクアレイシステムの構成を示すブロック図である。

【符号の説明】

1 ディスクアレイ装置

2 x (x = a、b、……、n) 計算機

3 経路制御装置

4 管理コンソール

5 x (x = a、b、……、n) チャンネルパス

6 x (x = a、b、……、n) LAN (Local Area Network)

7 通信手段

10 デバイスネットワーク

11 x (x = a、b、……、n) コントローラ

12 x (x = a、b、……、n) ディスク装置

20 テープ装置

31 パス変更手段

41 管理ユーティリティ

111 x (x = a、b、……、n) データキャッシュ

112 x (x = a、b、……、n) 構成管理手段

121 a ~ 121 n LU (Logical Unit)

100 ディスクアレイ装置

200 x (x = a …… n) コントローラ

300 共有キャッシュ

400 共有メモリ

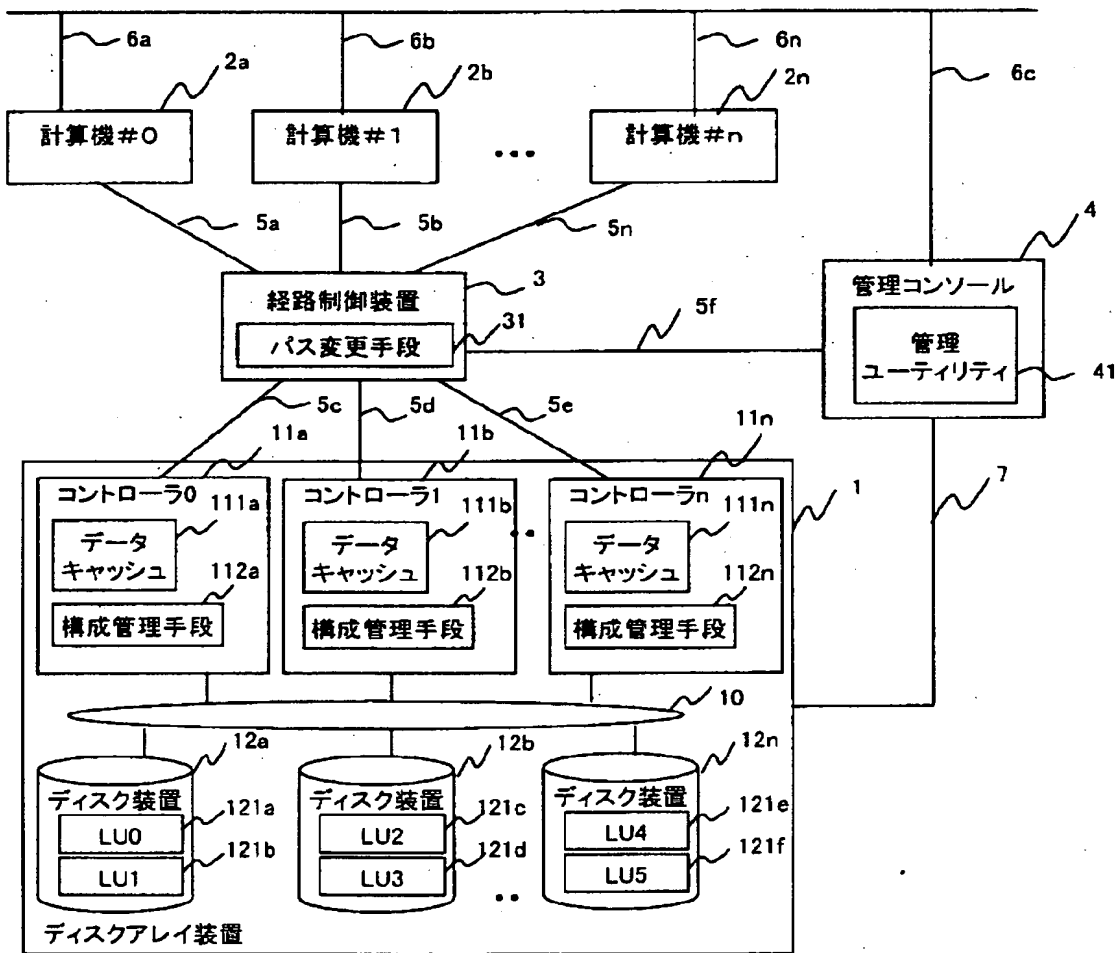
500 x (x = a …… n) ディスク装置

600 共通バス

【書類名】 図面

【図1】

図1



【図 2】

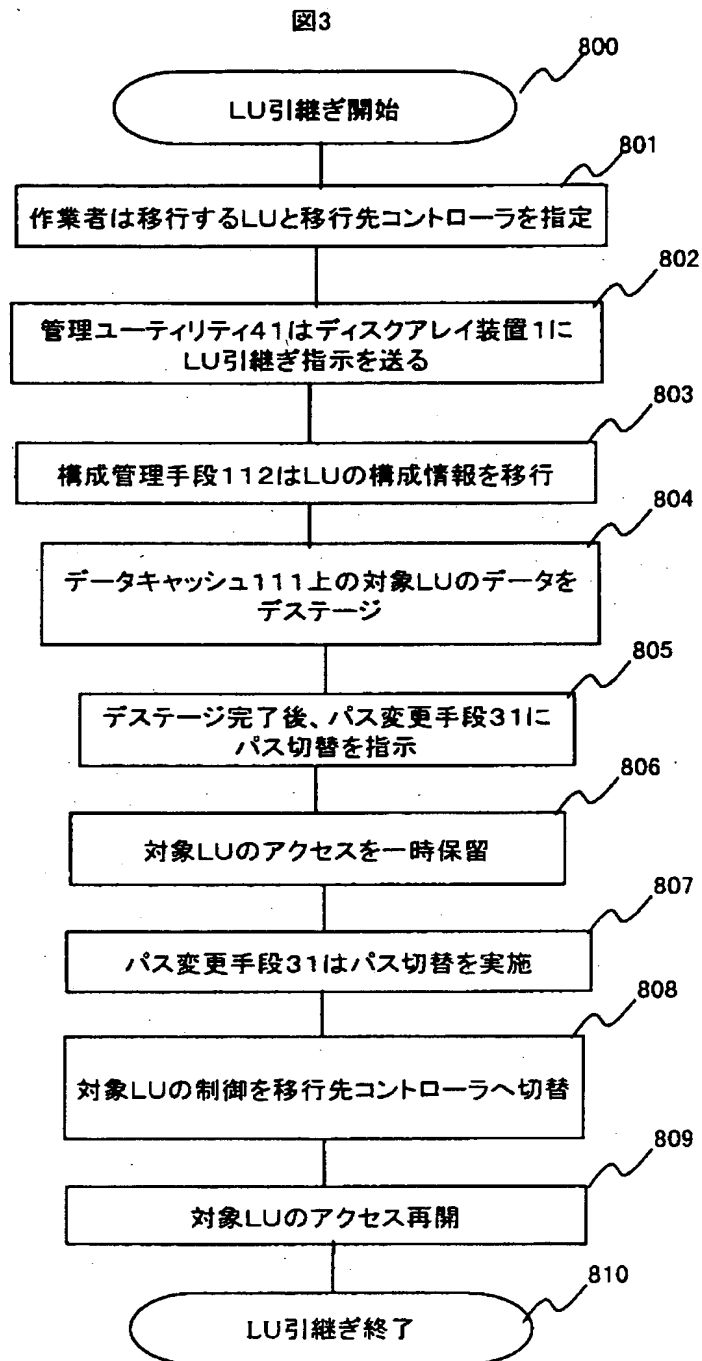
構成情報テーブル

図2

1121

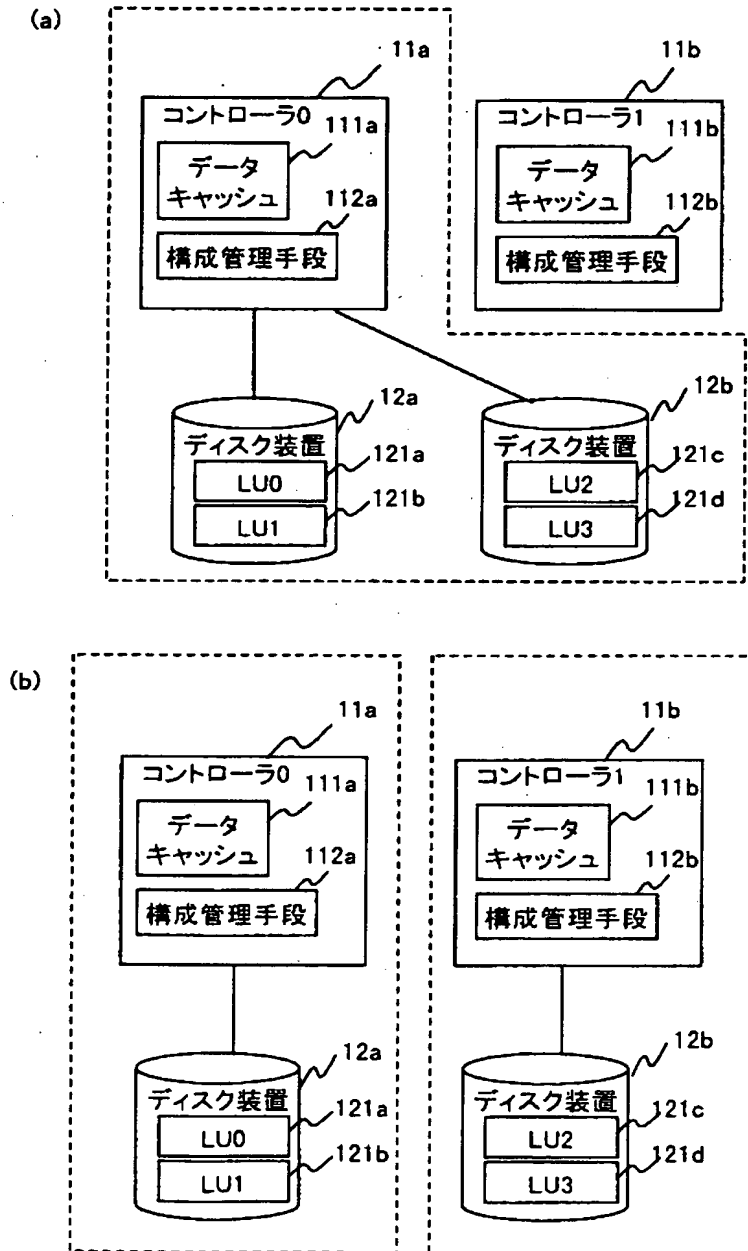
内部 LUN	外部 LUN	ポート 番号	コントローラ番号		ブロック数	RAIDGroup 番号	RAID レベル	物理アドレス 情報
			デフォルト	カレント				
...

【図 3】



【図4】

図4



【図 5】

図5

(a)

1121x

引継ぎ前の構成情報テーブル

内部 LUN	外部 LUN	ポート 番号	コントローラ番号		ブロック数	RAIDGroup 番号	RAID レベル	物理アドレス 情報
			デフォルト	カレント				
0	1	0	0	0
1	3	0	0	0
2	0	1	0	0
3	2	1	0	0

(b)

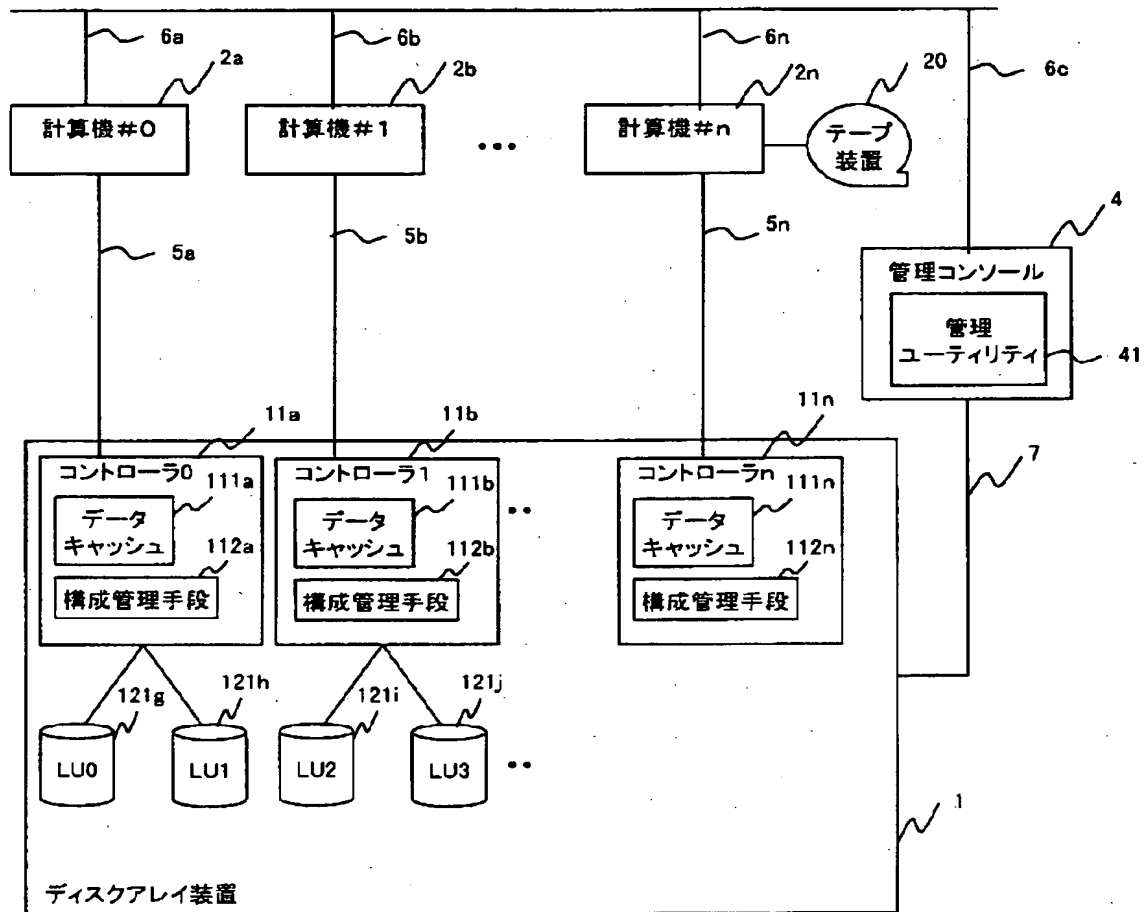
1121x

引継ぎ後の構成情報テーブル

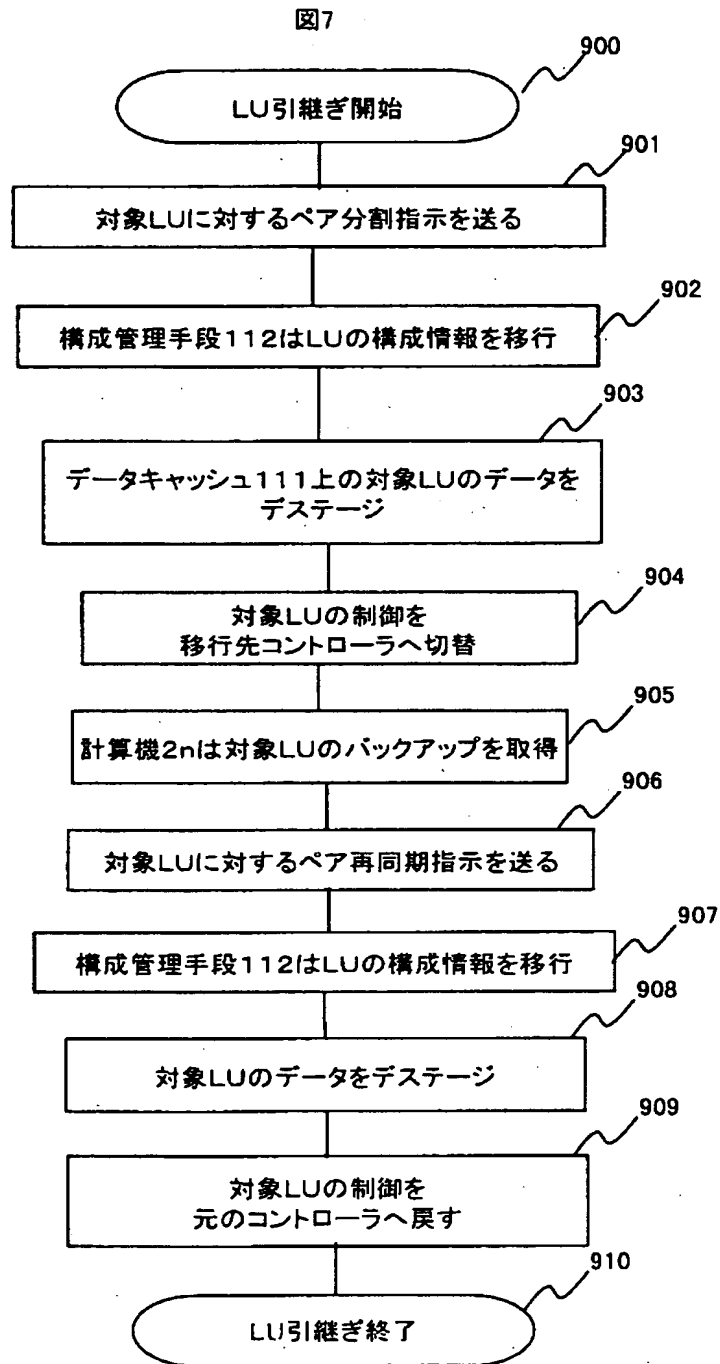
内部 LUN	外部 LUN	ポート 番号	コントローラ番号		ブロック数	RAIDGroup 番号	RAID レベル	物理アドレス 情報
			デフォルト	カレント				
0	1	0	0	0
1	3	0	0	0
2	0	0	1	1
3	2	0	1	1

【図 6】

図6



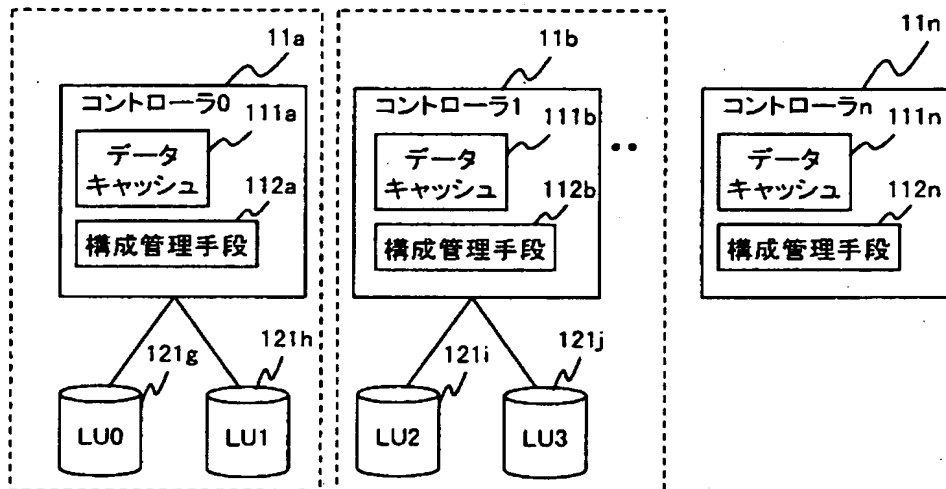
【図 7】



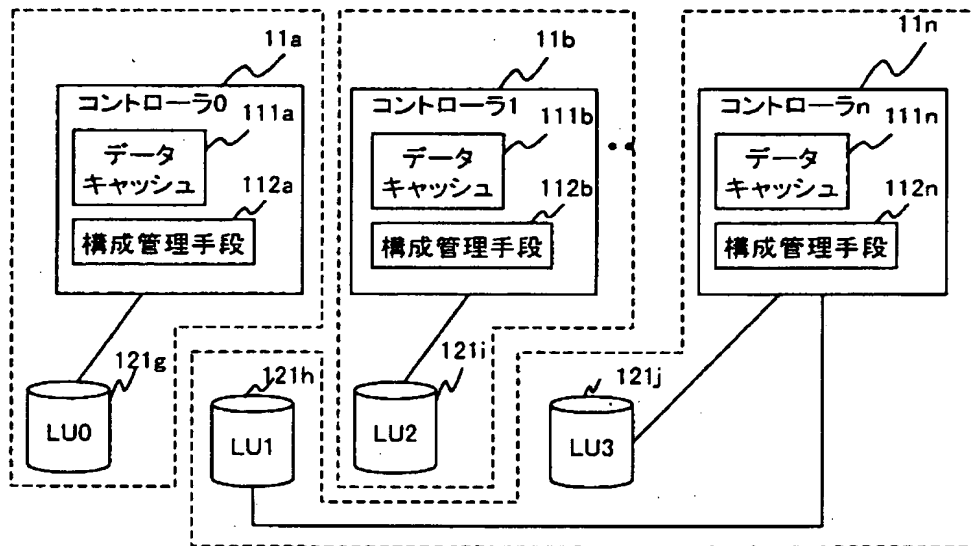
【図 8】

図8

(a)



(b)



【図9】

図9

(a)

1121x

引継ぎ前の構成情報テーブル

内部 LUN	外部 LUN	ポート 番号	コントローラ番号		ブロック数	RAIDGroup 番号	RAID レベル	物理アドレス 情報
			デフォルト	カレント				
0	0	0	0	0	1	1	1	1
1	1	0	N	0	1	1	1	1
2	1	0	1	1	1	1	1	1
3	2	0	N	1	1	1	1	1

(b)

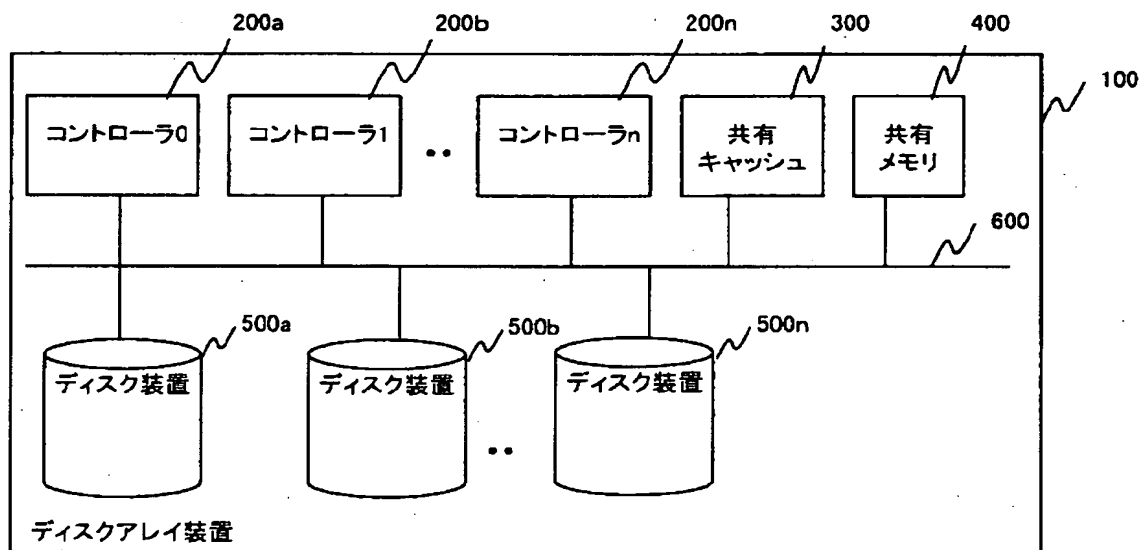
1121x

引継ぎ後の構成情報テーブル

内部 LUN	外部 LUN	ポート 番号	コントローラ番号		ブロック数	RAIDGroup 番号	RAID レベル	物理アドレス 情報
			デフォルト	カレント				
0	0	0	0	0	1	1	1	1
1	1	0	0	N	1	1	1	1
2	1	0	1	1	1	1	1	1
3	2	0	1	N	1	1	1	1

【図 1 0】

図10



【書類名】 要約書

【要約】

【課題】 システムを中断させることなく任意のコントローラ間で任意のボリュームを引き継ぐことを可能とした各コントローラが独立した専用キャッシュを備えたディスクアレイシステム。

【解決手段】 コントローラ相互間でボリュームを引き継ぐとき、引き継ぎ元コントローラは、データキャッシュ上の対象ボリュームのデータをディスクに掃き出してディスクの整合性を保つようにする。これにより、各コントローラが独立した専用キャッシュを備えた場合においてもシステムを中断させることなく任意のコントローラ間で任意のボリュームを引き継ぐことが可能となる。各コントローラは構成管理手段を備え、移行すべきコントローラ番号を記録しておくことにより、自動的なボリュームの引継ぎ及び元のコントローラに制御を戻すことができる。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000005108]

1. 変更年月日	1990年 8月31日
[変更理由]	新規登録
住 所	東京都千代田区神田駿河台4丁目6番地
氏 名	株式会社日立製作所